system is built on what we learned from the Peace Corps—how to work with slender resources."

Although the Peace Corps has for 19 years emphasized community-managed health care, recent efforts have strengthened Peace Corps' health-care focus. All 5,500 Peace Corps volunteers—whether working in fisheries, agricultural produce, small animal husbandry, appropriate technology, sanitation, women's clubs, village energy projects, or nutrition—are currently trained in fundamentals of individual and community managed health. In addition, the proportion of volunteers working primarily in the fields of health care and nutrition is steadily increasing—from 15 per cent in 1978 to 19 per cent in 1979. The latter figure represents over 1,000 individuals.

These skill-trained volunteers continually demonstrate that it is not necessary to be a physician to improve public health. Some say that the best thing ever to happen to the health of one coastal West African country was the popularity of a song written by a Peace Corps volunteer, "Keep the Flies Out of Your Baby's Eyes." Peace Corps volunteers organize and manage regional immunization teams, they train networks of health extension "agents", they organize nutrition surveillance campaigns to identify severe malnutrition, they develop rural health clinics, they set up community outreach programs, they train village health workers, and they augment the skills of midwives.

The volunteers essentially strengthen the ability of ordinary people to care for themselves and for each other.

The challenge posed by the goals of the Alma Ata conference leave us very little room—or time—for self-congratulations. Yet it should be encouraging, in the face of Russian-backed Cuban physician deployment to Third World countries for political purposes, that US volunteer efforts have followed the route of mutual cooperation to improve Third World health care, rather than imposing cadres of technical experts. The Peace Corps experience represents a step in the right direction—but we have many more steps to take if adequate health care for all people is to be a reality 20 years from now.

Criteria, Norms and Standards of Quality: What Do They Mean?

AVEDIS DONABEDIAN, MD, MPH

Abstract: Quality assessment requires specification of: 1) a set of phenomena that are usually attributes of either process or outcome; 2) a general rule of what constitutes goodness; and 3) a precise numerical statement of what constitutes acceptable or optimal goodness with respect to each of these phenomena. The terms "criteria," "norms," and "standards," as currently employed, do not correspond well with these three components, but they could be used effectively if the basic distinctions were understood. Alternatively, one could use, as corresponding terms, "elements," "parameters," and "standards." The terms "criteria," and "norms" would then be redefined and be available to be used more uniformly, while "standards" could be further differentiated according to method of measurement, configuration, level, and flexibility. (*Am J Public Health* 1981; 71:409-412.)

Everyone agrees that in order to assess the quality of medical care one needs "criteria," "norms," and "standards." Unfortunately, we have used these words in so many different ways that we no longer clearly understand each other when we say them. But we have used them for so long, that we do not have the liberty of abandoning them entirely, so as to begin all over again. Besides, what better new words would we find to say what we need to say. Our more reasonable course of action, therefore, is to see whether we can clarify the existing nomenclature, barnacled and misshapen though it may be with the encrustations of careless past usage.

Address reprint requests to Avedis Donabedian, MD, MPH, Nathan Sinai Distinguished Professor of Public Health, Department of Medical Care Organization, School of Public Health, University of Michigan, Ann Arbor, MI 48109. This paper was submitted to the Journal and accepted for publication on December 12, 1980.

Basic Elements of a Nomenclature

As a first step, it would perhaps be useful to put aside the words themselves, and to consider what thoughts we need to express in order to deal with the subject at hand. To do so, let us assume that we have agreed on what quality means, and have also agreed to examine either the process of care or its outcomes, as a means to its assessment. It seems

to me that we now need three different things. The first is a set of discrete, clearly definable, and precisely measurable phenomena that belong within the categories of process or of outcome, and that, in some specifiable way, are relevant to the definition of quality. These phenomena can be viewed as elements, components, attributes, or characteristics of either process or outcome. These things, whatever they are called, must be so clearly defined that we can say with confidence whether they are present or absent. It follows that we can measure them at least by saying how often they are present. Additionally, it may be possible to use a numerical measure of their quantity. For example, the taking of a blood pressure reading is an element of the process of care whose frequency can be reasonably well established, whereas the actual blood pressure reading is an element of the outcome of care for hypertensives whose magnitude can be easily determined.

Given these phenomena (elements, components, attributes, or characteristics) we next need some general rule as to what constitutes goodness with respect to each: for example, that their presence is better than their absence, or that a larger quantity is better than a smaller one. Obviously, this general rule derives from the definition of quality, and from the manner in which the phenomenon in question relates to that definition. To continue with my examples, we would claim that blood pressure measurements contribute to the quality of care, so that the more the better, and that a lowering of an abnormally elevated blood pressure is (by definition) a desirable outcome of care.

The third, and final, member of our list is a more precise, numerical statement of what constitutes acceptable or optimal goodness with regard to each of the phenomena under study. For example, we might agree that 90 per cent of all adult patients who see a physician for any reason ought to have had their blood pressure taken within a six-month period. We may also agree that of a group of younger adult patients with hypertension, 70 per cent should have a diastolic pressure of 90 mms or below within one year of the initiation of therapy.

I would like to call the phenomena that one counts or measures in order to assess the quality of care its "criteria." The general rules that indicate what is goodness I would like to call "norms." The precise count or quantity that specifies an adequate, acceptable, or optimal level of quality I would like to call a "standard." Obviously, current usage does not permit me this luxury. Nor am I certain that I could, myself, consistently observe in my own language the fine distinctions that this nomenclature would demand. In particular, the criteria of assessment often imply the general rules or norms that constitute goodness, so that the distinction between criterion and norm is often not worth making. But this is only a minor problem as compared to the resistance that the established variants of this nomenclature would pose.

Some Current Nomenclatures

"Criteria," "norms," and "standards" are, of course, words of long-established usage, endowed with a variety of meanings, often depending on the context in which they are used. But, of late, one set of meanings has been imposed upon them by administrative fiat, through the definitions that were adopted by the National Professional Standards Review Council, in consultation with the Task Force on Guidelines of Care of the American Medical Association Advisory Committee on PSRO.[1] The definitions appear in the *PSRO Program Manual* as follows:[2]

> "*Criteria*—Medical care criteria are predetermined elements against which aspects of the quality of medical service may be compared. They are developed by professionals relying on professional expertise and on the professional literature."
> "*Norms*—Medical care appraisal norms are numerical or statistical measures of usual observed performance."
> "*Standards*—Standards are professionally developed expressions of the range of acceptable variation from a norm or criterion."

It is, I believe, unfortunate that these definitions have become part of our officially endorsed and propagated language, for they tend to confound the basic distinctions that I tried to make in the preceding section. The term "criteria," as defined for use by the PSROs, is closest to my proposed usage of the word, as signifying those elements that are to be counted or otherwise measured in the process of quality assessment. But the "norms," as used in this official formulation, do not necessarily have normative force, since they signify neither goodness nor badness, but simply an observed phenomenon. The norm is, in fact, the average observed quantity of a phenomenon that *could* serve as a criterion. Hospital length of stay, for example, is a possible criterion of the quality, intensity, or efficiency of care. But the observed mean or median length of stay, the standard deviation of its frequency distribution, or the ranges of specified segments of that distribution, are all nothing more than measures that characterize that criterion. They have no normative or evaluative connotation, until one adds a specification of what is good or bad about these measurements.

In the offical nomenclature, the "norms" of practice are, strictly speaking, nothing more than its descriptors. Unfortunately, however, the term is not so neutrally or strictly used. There is always at least a covert implication that to conform to the average is a good, and to fall far outside it is either bad, or cause for suspicion that it may be bad. When used with this covert or overt meaning attached to it, the "norm" is a criterion to which a normative statement, based on average experience, has been added. If, as is often the case, a criterion is an element or attribute of either process or outcome to which an evaluative connotation based on professional opinion is overtly or covertly attached, the distinction between a criterion and a norm is reduced to this: that in the former the value judgment is derived from the opinions of the professionals, whereas in the latter the value judgment derives, at least in part, from an observation of their practice.

The "standards," as officially defined, seem to be a specification of the limits of tolerance for departures from the values attached to the criteria and the norms. This makes sense with respect to the "norms" if these are interpreted, in the strict sense, as having value-free quantitative characteristics. But the official definition of "standards" gives us the

first inkling that the "criteria" are not merely attributes of either process or outcome, but also come with measurements or values attached. Therefore one is led to the conclusion that the criteria come accompanied by their own standards, and that the "standards" of the official definition are the tolerable deviations from these other, preexisting, standards.

I do not think that I am engaged in willful misconstruction. I believe that the problem is genuine, and that it flows from insufficient attention to basic conceptual structure, as well as to language. Partly to prove my point, but mainly to acquaint the reader with an important alternative, I will now reproduce part of a nomenclature proposed by Slee that, I think, is a model of reasonableness and clarity. This may be because Slee disposes altogether with the pesky "criteria," and offers us, instead, the following vocabulary.[3]

> "*Parameter:* An objective, definable, and measurable characteristic of the patient himself or of the process or outcome of his care. Each parameter has a scale of possible values—for example, age in years; a drug given or not given, or the dosage; final outcome, death or life."
> "*Norm:* A statistical description of the central tendency of the observed values of a selected parameter, along with a measure of the variability of the values, taken from an adequate sample of corresponding studies . . ."
> "*Standard:* The *desired* achievable (rather than the *observed*) performance or value with regard to a given parameter."

In the nomenclature proposed and used by Slee, "parameter" corresponds to what I proposed to call "criterion," and "standard" is used in the same way that I would. We differ on the usage of "norm," but Slee, and for that matter the PSROs as well, are on firm ground, since "normal" is recognized to mean usual as well as good. However, it would be useful to avoid the confusion caused by this dual connotation, especially in an evaluative context that may constantly prompt misunderstanding. The difficulty is in finding another word, free of valuational connotations, to stand for "norm." I have none of my own to suggest, but it may be that the vocabulary offered by Slee does.

The key to one possible solution may be in the word "parameter," which appears to have a variety of technical meanings depending on the branch of science in which it is used. Speaking as statisticians, Hagood and Price[4] distinguish a "parameter" and a "statistic" as follows: "The universe value of any summarizing measure of a distribution of one or more characteristics is called a parameter; the value of a summarizing measure observed in a sample is called a statistic." It seems to me, therefore, that what Slee calls a "norm" could readily be called a "parameter;" whereas what he calls a parameter could easily be called an element, component, attribute, or variable, assuming one wanted to avoid, as I know Slee would, the use-encrusted "criterion." Elements, parameters, standards: it is a nomenclature that might possibly work. In the process, it would liberate the word "norm," so it can be used with less specificity to mean a general rule of what constitutes goodness. And the word "criterion" would also remain as a usefully flexible, although imprecise, term that would mean an element or attribute that is to be used in evaluation, and that often comes accompanied by an explicit or implicit norm (as I would use the word) or, even, by a standard. Most lists of "criteria" are, in fact, composed of such items, since what the lists say is that these items should be found in the care of all or most patients with a specified diagnosis.

It is in this inclusive sense that the word "criterion" appears in the lexicon of the Performance Evaluation Procedure (PEP) that was developed under the auspices of the Joint Commission on Accreditation of Hospitals. As Jacobs, *et al,* put it, "A criterion in the PEP system consists of an element of major consequence, a standard, and exceptions to the standard, if any. The audit committee decides which *elements* are of such importance to patient outcome that, if unmet, review of the chart is warranted. To each element is added a *standard,* which is always set at 100 per cent or 0 per cent to facilitate the screening of charts." As to the *exceptions,* these specify the circumstances under which departures from the standards are probably justifiable and, therefore, which allow the record to go through without detailed review.[5]

Some Varieties of Standards

The nomenclature described by Jacobs, *et al,* suggests that the tripartite structure that I suggested earlier in this paper can easily be reduced to two: "elements," and "standards." Obviously, the "norms," as I would call them, are included in the "standards," because these latter are simply more specific statements of the norms.

The standards may, themselves, be expressed in a variety of forms, depending in part on the way in which the elements being evaluated are measured, and, in part, on the nature of the relationship between what is being measured and what is defined as good.

Certain elements of care or of its outcomes are measured in a nominal form, as present or absent. The standards for these are stated as a per cent of cases in which the element does or does not occur. Since this is such a frequent form, Slee has given it the distinctive name of "pattern standard." This he defines as "A percentage indicating how often a given outcome or process parameter value occurs per 100 patients if the care given is excellent. For example, death rate in acute myocardial infarction, normal tissue rate in appendectomy, performance of urinalysis for all patients."[3] Slee does not offer a term to signify a standard that applies to variables that are capable of being measured with a more developed numerical scale. Examples of these are the length of hospital stay, or the blood pressure reading. For such variables, scalar standards that specify mean values and tolerable deviations from such values are, of course, possible.

The relationship between the element being measured and the judgment of what is good gives rise to some interesting variants of standards. There are certain things of which the more we have the better, or the less we have the better. Obvious examples are survival, on the one hand, and death, on the other. The norms and standards that pertain to such phenomena may be spoken of as "monotonic." We have already seen, in the PEP method of assessment, an at-

tempt to force all standards into the two polar positions on this continuum, by requiring that they be specified as 0 per cent or 100 per cent. Perhaps such all-or-nothing standards could be called "categorical." By contrast, there are phenomena which have their most desirable value at some maximum or minimum point, on either side of which the valuation placed upon them is less. The norms and standards that pertain to such things may be called "inflected." The frequency of surgery for suspected appendicitis is an example. There is an optimal propensity to operate, the results being worse if the surgeon is either too ready or too reluctant to operate.[6]

The standards can, of course, be set at a variety of levels, depending on the level of quality that one wishes to attain, or to use as a benchmark for comparison. Accordingly, a variety of terms have come into use that refer, directly or indirectly, to the level of the standard. For example, Williamson speaks of "maximum conceivable benefit," and of "achievable benefit."[7] Slee ties "excellent care" to a "realistically achievable" standard; but, as I see it, he allows the level of the standard to vary almost at will by saying, in effect, that the standard can be any "desired achievable" value.[3]

Sometimes the standard is specified as a single value. Sometimes a certain degree of "tolerance" is introduced by specifying a threshold below which the care for a group of patients must fall before an institutional response is considered to be necessary. For example, the system of monitoring described by Slee[3] provides for a toleranced standard, with threshold values which must be crossed before a "signal" that calls for action is considered to have sounded.

When the normative structure underlying the standard is monotonic, only one threshold is needed, either below or above the target standard, depending on whether the highest or lowest point of the range of possible values represents the best performance. With an inflected normative structure, the standard, whether it is the best possible or merely the desired value, lies within a range that is bounded by a threshold on either side. To borrow from the nomenclature of industrial quality control, these might be called the lower and upper control levels or limits, respectively.[8]

Conclusions and Prospects

I fear that this paper may seem to be nothing more than a nit-picking exercise in semantics, especially since I am not sure that I, myself, would be able to always walk the narrow path of linguistic purity. The older words remain seductively attractive. Their very imprecision endows them with rich resonances that embellish both speech and writing. No doubt, we shall continue to use them. But I hope that the variants of meaning of which they are capable will have become clearer, now that we have examined the underlying conceptual structure to which they give voice. And, although we may still continue to talk in many tongues, I hope that, from now on, we shall be able to understand what we intend to say to one another.

ACKNOWLEDGMENTS

This paper is based on a section in a forthcoming book by Professor Donabedian on *The Criteria and Standards of Quality* to be published by the Health Administration Press, Ann Arbor. Work on the book was supported by the Commonwealth Fund, and by the National Center for Health Services Research under Grant No. HS-02081. Copyright to the text is held by the Regents of the University of Michigan, who have kindly permitted publication in this form. The opinions expressed in the paper are those of the author, and they do not represent any of his sponsors.

REFERENCES

1. American Medical Association Advisory Committee on PSRO, Task Force on Guidelines of Care: PSROs and norms of care. JAMA 1974; 229:166–171.
2. Department of Health, Education, and Welfare, Office of Professional Standards Review: PSRO Program Manual, chapter VII, p 16, March 15, 1974.
3. Slee V: PSRO and the hospital's quality control. Annals of Internal Medicine 1974; 81:97–106.
4. Hagood MJ, Price DO: Statistics for Sociologists. New York: Henry Holt, 1952, p 220
5. Jacobs CM, Christoffel TH, Dixon N: Measuring the Quality of Patient Care: The Rationale for Outcome Audit. Cambridge, MA: Ballinger, 1976, pp 51–52 and ff.
6. Neutra R: Indications for the surgical treatment of suspected acute appendicitis: a cost-effectiveness approach. Chapter 18, pp 277–307, IN: Bunker JP, Barnes BA, Mosteller F, (eds): Costs, Risks, and Benefits of Surgery. New York: Oxford University Press, 1977.
7. Williamson JW: Assessing and Improving Health Care Outcomes: The Health Accounting Approach to Quality Assurance. Cambridge, MA: Ballinger, 1978, p 61.
8. Fetter RB: The Quality Control System. Homewood, IL: Richard D. Irwin, 1967, pp 38 ff.